

## **L'Item Response Theory come strumento di valutazione delle eccellenze nella scuola**

Cristiana Ceccatelli<sup>1</sup> Tonio Di Battista<sup>2</sup>  
Francesca Fortuna<sup>3</sup> Fabrizio Maturo<sup>4</sup>

**Sunto.** Nella prospettiva delle moderne teorie del capitale umano, il processo formativo acquista un ruolo centrale per lo sviluppo di un paese. Incoraggiare un apprendimento attivo e consapevole diventa, quindi, fondamentale. Questo lavoro presenta il caso delle Olimpiadi Nazionali di Statistica, un'iniziativa volta alla promozione del ragionamento statistico nella Scuola Secondaria di Secondo Grado. In particolare si analizza uno dei questionari utilizzati durante la competizione ricorrendo ai modelli dell'*Item Response Theory* con l'obiettivo di evidenziarne possibili criticità e di sottolineare l'importanza dell'utilizzo di metodi statistici nel processo di valutazione.

**Parole Chiave:** Item Response Theory, abilità latenti, valutazione, capitale umano.

**Abstract.** According to the modern human capital theory, the learning process is assuming a central role for a country development. Therefore, it is very important to encourage an active and conscious learning. This paper shows the case of the National Olympic of Statistics, that is an initiative to improve statistic reasoning in the Second Grade Middle School. In particular, it has been analyzed, through Item Response Theory models, only one of the competition questionnaires. The aim of this work is to provide a critical review of the questionnaire and to highlight the importance of using statistical methods in evaluation processes.

---

<sup>1</sup> c.ceccatelli@unich.it

<sup>2</sup> dibattis@unich.it

<sup>3</sup> francesca.fortuna@unich.it

<sup>4</sup> f.maturo@unich.it

**Keyword:** Item Response Theory, latent ability, evaluation, human capital.

## **1. Innovazioni didattiche per lo sviluppo del capitale umano in istruzione**

Il capitale umano rappresenta l'insieme di conoscenze, competenze, abilità, emozioni, acquisite durante la vita da un individuo e finalizzate al raggiungimento di obiettivi sociali ed economici, singoli o collettivi [2]. *Formazione, capitale umano* ed *e-learning organization* sono capitoli importanti della strada che oggi tutti i capitalismi nazionali stanno intraprendendo, affidandosi non solo alle innovazioni di singole imprese ma anche a forme di intelligenza collettiva [7].

L'educazione diventa, quindi, un fattore importante per sostenere la crescita economica di un paese e, in questo contesto, si pone il problema di ripensarla soprattutto a livello scolastico. Tradizionalmente, infatti, la trasmissione della conoscenza è centrata sull'idea di un processo unidirezionale, ingabbiato in strategie e metodologie statiche di trasmissione basate sull'ascolto, sull'imitazione e sulla riproduzione mnemonica degli apprendimenti.

Negli ultimi anni, la ricerca, in campo psicologico, pedagogico e didattico, ha messo in evidenza come non sia sufficiente trasmettere informazioni, saperi e nozioni, per promuovere la conoscenza [11]. Quest'ultima, infatti, non è insita nei contenuti del sapere, come insieme di dati strutturati e formali, ma è necessario che il soggetto che apprende sia motivato e desideroso di allargare i propri orizzonti del sapere. Il processo di insegnamento-apprendimento deve, quindi, esplicitarsi in un clima sociale fertile che assicuri uno scambio ricco e dinamico di conoscenza. In questo scenario diventa fondamentale promuovere eventi che coinvolgano attivamente e consapevolmente gli studenti stimolando in questi il desiderio della scoperta e del confronto [3]. Le Olimpiadi Nazionali di Statistica sono un esempio di tali iniziative, rappresentando, quindi, un'opportunità concreta di investimento nel campo dell'istruzione. Un investimento che mira al

potenziamento di un atteggiamento positivo e propositivo dei giovani nei confronti della conoscenza, facendo leva sulle nuove tecnologie dell'informazione e della comunicazione con un'attenzione particolare alla disciplina statistica.

Le Olimpiadi di Statistica, promosse per la prima volta dalla SIS (Società Italiana di Statistica) nel 2011, hanno riscosso immediatamente un grande successo che è andato crescendo nel corso degli anni, come dimostrato dal numero dei partecipanti che è passato da 578 del 2011 a 1139 del 2012. Gli obiettivi dell'iniziativa erano molteplici ed attenevano principalmente ad avvicinare gli studenti al ragionamento logico e statistico; promuovere una lettura funzionale dei fenomeni reali osservati; incentivare l'uso delle tecnologie informatiche per fini formativi e diffondere best-practices didattiche. La competizione ha previsto l'iscrizione facoltativa da parte delle Scuole Secondarie di Secondo Grado, le quali hanno proposto le eccellenze all'interno delle varie classi IV e V. In ragione della libera adesione all'iniziativa, la dislocazione territoriale delle scuole partecipanti sul territorio italiano e la numerosità degli studenti all'interno delle classi non sono omogenee.

La seconda edizione delle Olimpiadi nazionali di Statistica si è svolta nei giorni 1 e 2 Marzo 2012 ed è stata strutturata in due categorie a seconda che l'insegnamento della statistica fosse previsto all'interno di corsi specifici (categoria statistica) oppure all'interno di programmi di matematica (categoria matematica). Per ciascuna categoria e classe si è svolta una fase eliminatoria alla quale hanno partecipato tutti gli iscritti ed una fase finale rivolta esclusivamente ai migliori. Per ciascuna delle due fasi e per ciascuna categoria, è stato redatto un questionario a risposta multipla composto da 20 domande per la fase eliminatorie e 15 per la finale. Negli otto diversi questionari sono stati affrontati argomenti di statistica descrittiva, probabilità, principi inferenziali, matematica e logica. L'accesso alla finale è stato regolamentato sulla base dei seguenti criteri di ammissione:

- 1) punteggio minimo di 20/60;
- 2) miglior punteggio;
- 3) miglior tempo;

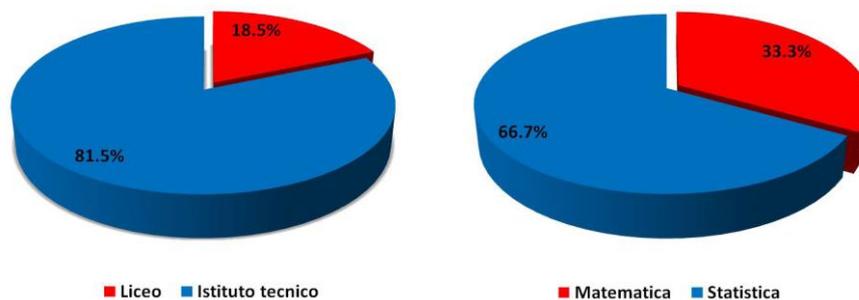
- 4) un minimo di 1 studente per classi fino a 20 alunni; procedendo per multipli di 10.

La gara si è svolta nelle aule informatiche delle singole scuole alla presenza di un tutor della scuola e di un tutor esterno per garantire il corretto svolgimento della prova. In contemporanea, da tutte le scuole d'Italia, i partecipanti hanno effettuato l'accesso ad una piattaforma e-learning Moodle gestita dal server del CINECA in collaborazione con l'Università "G. d'Annunzio" di Chieti-Pescara.

## 2. La competizione in cifre

Alla competizione hanno preso parte 54 scuole, rappresentate per il 18.5% da licei scientifici e per l'81.5% da istituti tecnici industriali (Fig.1). Data la predominanza di ITI si denota una corrispondente prevalenza della categoria statistica (66.7%) rispetto a quella di matematica (33.3%), in quanto nei licei scientifici, generalmente, la statistica è studiata all'interno del corso di matematica.

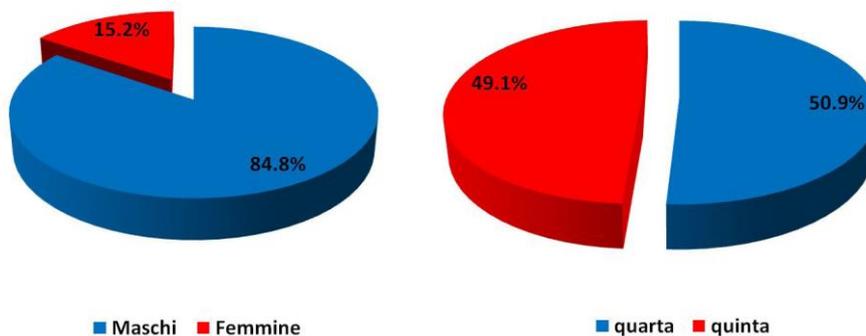
Fig. 1: Distribuzione delle scuole partecipanti per tipologia di istituto e categoria di corso



Il numero totale di studenti in gara è stato pari a 1139 con una predominanza di maschi (84.8%) rispetto alle femmine (15.2%). Per quanto riguarda, invece, le classi partecipanti, il collettivo appare distribuito uniformemente tra le quarte e le quinte (Fig. 2).

Rispetto alla ripartizione territoriale, si nota una prevalenza di studenti del Sud (46.3%) e del Nord (40.7%) a discapito di quelli del Centro (13.0%) (Fig. 3).

Fig. 2: Distribuzione dei partecipanti per genere e per classi



In particolare si riscontra una maggiore concentrazione di iscritti in Sicilia, Puglia, Campania, Lombardia e Veneto (Fig. 4); mentre alcune regioni non hanno affatto preso parte alla competizione (Molise, Basilicata, Valle D'Aosta ed Umbria).

Fig. 3: Distribuzione dei partecipanti per area territoriale

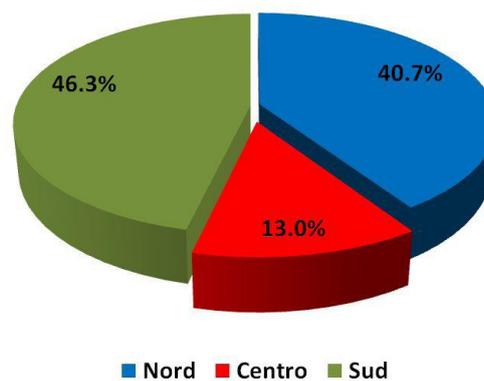
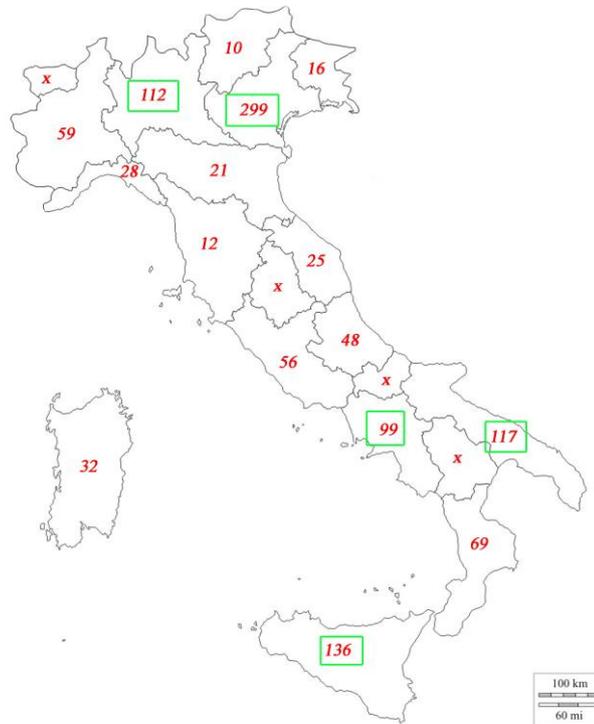


Fig. 4: Distribuzione dei partecipanti per regione (valori assoluti)



### 3. Analisi del fenomeno attraverso l'Item Response Theory

L'obiettivo principale delle Olimpiadi di Statistica rimane quello della valutazione delle abilità degli studenti. Tradizionalmente lo studente viene valutato sulla base del punteggio grezzo ottenuto ad un test senza considerare altre informazioni quali ad esempio la difficoltà del test stesso. L'abilità, in realtà, rappresenta un concetto latente, ovvero un costrutto non direttamente osservabile e che, nella concretezza, viene misurato attraverso indicatori parziali, cioè variabili osservabili [5]. Nel caso in questione, le variabili osservabili sono le risposte fornite alle diverse domande (item) del questionario e, soltanto analizzando gli item, sia singolarmente che a livello aggregato, è possibile ottenere una valutazione più attendibile degli

studenti. Appare evidente, quindi, la necessità di definire, sulla base dell'insieme di domande osservabili, una misura sintetica e quanto più oggettiva dell'abilità latente.

In questo contesto si inseriscono i modelli dell'Item Response Theory (IRT) che sono volti a stimare la probabilità di rispondere correttamente ad ogni singolo item, in funzione del livello di abilità posseduto dal soggetto e di alcuni parametri propri degli item [1]. Secondo questo approccio è possibile prendere in considerazione tre parametri riguardanti gli item: la difficoltà, il potere discriminante, cioè la capacità di distinguere soggetti con diversi livelli di abilità ed il guessing, cioè la probabilità di rispondere correttamente per un soggetto con abilità molto basse. A seconda del numero di parametri considerati, si distinguono diversi modelli IRT [1]: modello logistico a uno, due e tre parametri rispettivamente. A prescindere dal modello scelto, le assunzioni fondamentali sono:

- 1) unidimensionalità: gli item individuano un solo tipo di abilità latente;
- 2) indipendenza locale degli item: dato un certo livello di abilità, le risposte ad ogni item sono indipendenti tra loro.

In questo lavoro, l'analisi dei dati si focalizza su uno dei questionari utilizzati durante la seconda edizione delle Olimpiadi Nazionali di Statistica. La motivazione di tale scelta risiede nel fatto che i questionari delle finali sono stati tralasciati per l'esiguo numero di partecipanti; mentre per quanto concerne i test della fase eliminatorie si è scelto di esaminare una determinata categoria e classe per assicurare la confrontabilità dei risultati. In particolare si fa riferimento al questionario della classe V, categoria statistica, per un totale di 426 studenti e 20 item. Quest'ultimi, che in origine contemplavano una penalità per le risposte errate, sono stati ricodificati e resi dicotomici prevedendo, quindi, solo due alternative: risposta corretta o errata.

Obiettivo dell'analisi è quello di verificare la validità del questionario, ovvero esaminare se lo stesso è in grado di misurare esattamente ciò che si intende misurare [10]. A tal fine si è scelto di adottare il modello IRT a due parametri, la cui relazione matematica è espressa dall'equazione [8]:

$$P(X_{ij} = 1 | \theta, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad (1)$$

dove:

- $X_{ij}$  è la risposta del  $j$ -esimo soggetto per l'  $i$ -esimo item;
- $\theta_j$  è l'abilità del  $j$ -esimo soggetto;
- $a_i$  è il parametro di discriminazione per l'  $i$ -esimo item;
- $b_i$  è il parametro di difficoltà per l'  $i$ -esimo item.

La scelta del modello a due parametri è stata dettata dal confronto con il modello ad un parametro rispetto al quale mostra un miglior adattamento ai dati sulla base del criterio di informazione di Akaike (AIC) e del criterio di informazione Bayesiano (BIC) [4].

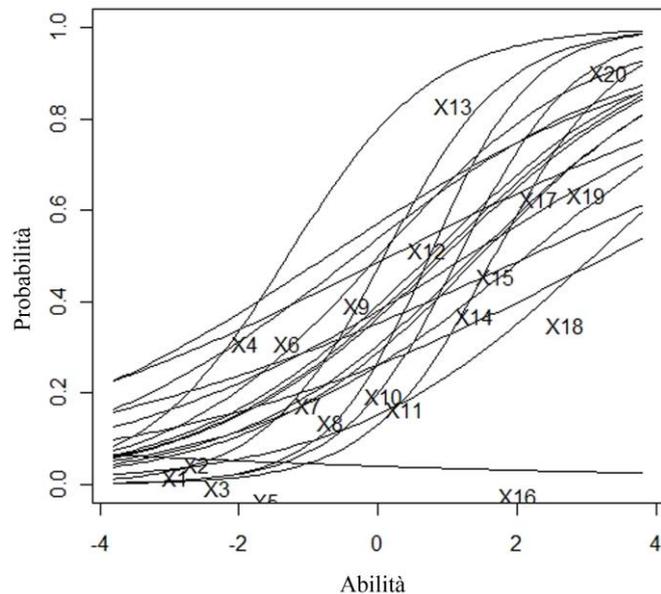
Tab. 1: Valori dei parametri per ogni item del questionario

<b>Item</b>	<b>Difficoltà (<math>a_i</math>)</b>	<b>Potere discriminante (<math>b_i</math>)</b>	<b>P(x=1 z=0)</b>
<b>13</b>	-1.30	0.97	0.78
<b>4</b>	-0.71	0.40	0.57
<b>6</b>	-0.32	0.47	0.54
<b>12</b>	0.01	0.67	0.50
<b>9</b>	0.19	0.31	0.49
<b>3</b>	0.10	1.15	0.47
<b>17</b>	0.74	0.59	0.39
<b>15</b>	1.30	0.38	0.38
<b>7</b>	0.87	0.59	0.37
<b>2</b>	0.95	0.59	0.36
<b>14</b>	2.19	0.28	0.35
<b>19</b>	1.39	0.59	0.30
<b>8</b>	1.46	0.62	0.29
<b>20</b>	0.74	1.36	0.27
<b>10</b>	3.33	0.31	0.26
<b>1</b>	2.13	0.49	0.26
<b>11</b>	1.17	1.19	0.20
<b>18</b>	3.10	0.56	0.15
<b>5</b>	1.67	1.14	0.13
<b>16</b>	-23.83	-0.13	0.04

La tabella 1 mostra i valori dei parametri  $b_i$  e  $a_i$  per ogni item e la probabilità di rispondere correttamente per uno studente medio. La stima dei parametri è stata ottenuta usando il criterio di massima verosimiglianza marginale con la quadratura di Gauss-Hermite [6], avvalendosi del software statistico R ed, in particolare, del package latent trait models (ltm) [9].

Le domande più difficili risultano essere la numero 10, 18 e 14; mentre le più semplici la 16, 13, 4 e 6. Per quanto riguarda il potere discriminante, sebbene esso possa variare da  $-\infty$  a  $+\infty$ , nella pratica, valori tra 0.8 e 2.5 sono considerati buoni. Nel nostro caso, gli item più discriminanti sono il numero 20, 11, 3, 5 e 13. In particolare è da segnalare la domanda 16 che risulta avere il peggior potere discriminante, tanto da assumere un valore negativo a significare che, studenti con bassa abilità, hanno maggiore probabilità di rispondere correttamente rispetto a studenti con media ed alta abilità. Alla luce di questa contraddizione appare evidente come tale item sia da eliminare dal questionario.

Fig. 5: Curva caratteristica (ICC) per ogni item



La figura 5 mostra la curva caratteristica (ICC) di ciascun item, cioè la funzione che esprime la probabilità di risposta all' *i*-esimo item in funzione dell'abilità. Differenti pendenze delle curve indicano differenti poteri discriminanti, mentre, con riferimento alla difficoltà e per un dato livello di abilità, si ha che le ICC degli item più difficili si trovano più in basso. Come conseguenza di queste due circostanze, nel modello a due parametri, è possibile stabilire un ordinamento degli item in base alla loro difficoltà solo in funzione di un prefissato livello di abilità e non in assoluto.

In figura 6 vengono rappresentati le ICC degli item che assumono valori estremi per i parametri di difficoltà e potere discriminante della tabella 1. Alla luce di quanto detto si può notare che l'item 20 è il più discriminante avendo una pendenza maggiore, mentre l'item 16, anche graficamente, appare un item da eliminare. Per quanto riguarda la difficoltà, a seguito dell'introduzione del parametro di discriminatività, si ha che un item possa essere considerato più o meno facile di un altro a seconda del livello di abilità che si considera. Ad esempio, l'item 20 risulta essere più difficile del 10 per livelli di abilità inferiori alla media, mentre risulta essere più facile nel caso opposto.

Fig. 6: Curva caratteristica (ICC) per gli item caratteristici

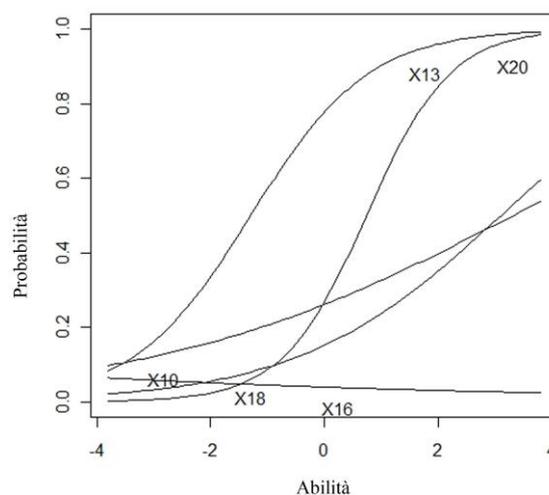
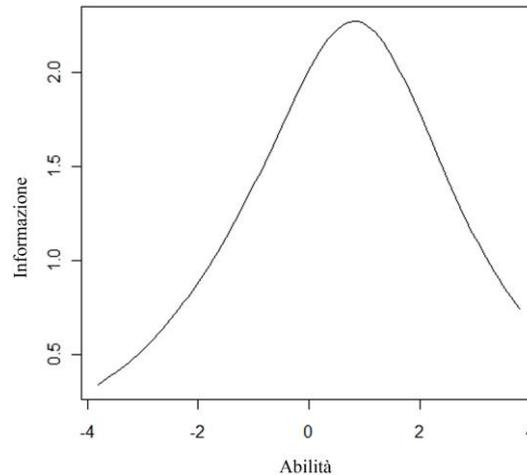


Fig. 7: Funzione di informazione del questionario



La figura 7 rappresenta la funzione di informazione dell'intero questionario nel modello a due parametri, ovvero il grado di precisione con il quale il test misura i diversi livelli di abilità. Formalmente la funzione di informazione è data dalla seguente equazione [1]:

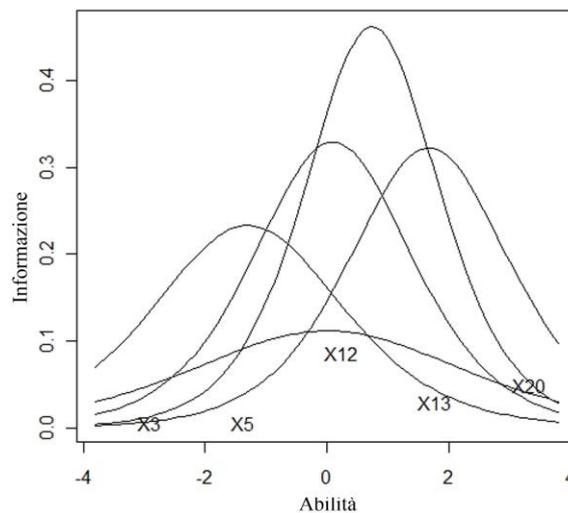
$$I(\theta) = \sum_{i=1}^n a_i^2 P_i(\theta) Q_i(\theta) \quad (2)$$

dove  $P_i(\theta) = 1 / \{1 + \exp[-a_i(\theta - b_i)]\}$  e  $Q_i(\theta) = 1 - P_i(\theta)$ . La figura evidenzia come, in media, il test misuri discretamente livelli di abilità pari a 1, ossia poco superiori alla media; questo è un risultato accettabile, ma, considerato che l'obiettivo principale della competizione è quello di misurare le eccellenze, sarebbe maggiormente auspicabile un questionario che discrimini meglio livelli di abilità più alti. Questo scopo è perseguibile analizzando le curve di informazione dei singoli item ed eliminando dal test le domande che diano un livello di informazione basso per le abilità alte o medio-alte. In particolare, in figura 8 sono rappresentati, per esigenze di chiarezza grafica, solo gli item che forniscono un maggior contributo informativo.

Tra questi, l'item che risulta maggiormente informativo per le abilità maggiori (Fig. 8) è il numero 5; infatti esso raggiunge il picco di contributo informativo (circa 0.3) in corrispondenza del livello di abilità 2; mentre tutti gli altri danno un buon livello di informazione solo per livelli di abilità inferiori. L'item 20, ad esempio, anche fornendo un buon livello di informazione, si presta maggiormente a valutare livelli di abilità inferiori all'item 5.

Dall'analisi delle curve di informazione, come già sottolineato, è possibile rimuovere dal test sia gli item con basso contributo informativo, sia quelli che danno un buon contributo ma solo per basse abilità, considerando che queste sono poco rilevanti nel contesto in esame. Nello specifico, sulla base dell'analisi effettuata, sarebbe altresì auspicabile inserire nel questionario alcuni item in grado di discriminare, con un buon livello di precisione, soprattutto abilità molto alte, ossia per livelli di  $\theta_j$  superiori a 2.

Fig. 8: Curva di informazione (ICC) per gli item caratteristici



#### 4. Considerazioni finali

Dall'analisi effettuata in questo lavoro emergono alcune criticità che mettono in evidenza l'importanza che riveste la calibrazione di un questionario per la valutazione delle abilità degli studenti.

La calibrazione, infatti, deve essere contestualizzata in relazione alla funzione valutativa che si intende operare [10]. Nel caso in cui l'obiettivo sia quello di valutare gli apprendimenti in ambito scolastico, sarebbe opportuno costruire un questionario che colga tutti i possibili livelli di abilità, considerando anche le situazioni estreme relative ai bisogni educativi speciali. Di contro, se la finalità è quella di selezionare i migliori all'interno di un *élite*, occorre predisporre strumenti di indagine, che siano in grado di discriminare i diversi livelli di abilità tra le eccellenze.

Alla luce di quanto detto, si evidenzia la necessità di un approccio statistico alla valutazione delle abilità che si espliciti attraverso un'interazione tra le competenze metodologiche dello statistico e quelle didattico-educative del docente.

#### Bibliografia

- [1] Baker F.B. (2001) *The basics of item response theory*, ERIC Clearinghouse on Assessment and Evaluation, Washington
- [2] Becker G. S. (1964) *Human capital*, University Press for the National Bureau of Economic Research, New York: Columbia
- [3] Ceccatelli C. (to appear, 2013) *La centralità della valutazione nella valorizzazione dell'Istruzione del Capitale Umano*, Franco Angeli, Milano
- [4] Ceccatelli C., Di Battista T., Fortuna F., Maturo F. (to appear, 2013) Best practices to improve the learning of statistics: the case of the National Olympics of Statistic in Italy, *Procedia - Social and behavioral Sciences*
- [5] Di Battista T. (2012) *Metodi e tecniche della valutazione: un approccio statistico*, Franco Angeli, Milano

- [6] Johnson M.S. (2007) Marginal maximum likelihood estimation of item response models in R, *Journal of Statistical Software*, 20 (10), 9-11
- [7] Levy P. (1996) *L'intelligenza collettiva. Per un'antropologia del cyberspazio*, Feltrinelli, Milano
- [8] Reckase M.D. (2009) *Multidimensional item response theory*, Springer, New York
- [9] Rizopoulos D. (2006) ltm: An R package for latent variable modeling and item response theory analyses, *Journal of Statistical Software*, 17 (5), 1-25
- [10] Van der Walt J.L. & Steyn, F. (2008) The validation of language tests. *Stellenbosch Papers in Linguistics*, 38, 191-204
- [11] Vertecchi, B. (2003) *Manuale della valutazione. Analisi degli apprendimenti e dei contesti, volume 4*, Franco Angeli, Milano